Low Frequency Interpolation of Room Impulse Responses Using Compressed Sensing

Remi Mignot, Gilles Chardon, Laurent Daudet

Low Frequency Interpolation of Room Impulse Responses using Compressed Sensing

Rémi Mignot, Gilles Chardon, Laurent Daudet

***Abstract*—Measuring the Room Impulse Responses within a finite 3D spatial domain can require a very large number of measurements with standard uniform sampling. In this paper, we show that, at low frequencies, this sampling can be done with significantly less measurements, using some modal properties of the room. At a given temporal frequency, a plane wave approximation of the acoustic field leads to a sparse approximation, and therefore a compressed sensing framework can be used for its acquisition. This paper describes three different sparse models that can be constructed, and the corresponding estimation algorithms: two models that exploit the structured sparsity of the soundfield, with projections of the modes onto plane waves sharing the same wavenumber, and one that computes a sparse decomposition on a dictionary of independent plane waves with time / space variable separation. These models are compared numerically and experimentally, with an array of 120 microphones irregularly placed within a $2\times2\times2$ m volume inside a room, with an approximate uniform distribution. One of the most challenging part is the design of estimation algorithms whose computational complexity remains tractable.**

***Index Terms*—Compressed Sensing, Room Impulse Responses, Wavefield reconstruction, Plane waves, Interpolation, Sparsity.**

I. Introduction

ACOUSTIC properties of a reverberating room can be given by analyzing its Room Impulse Responses (RIRs), which describe the acoustic transfer between sources and receivers. In [1], the concept of *Plenacoustic Function* (PAF) is introduced. This function gathers all RIRs of the room, and therefore it depends on time, on the source position, on the receiver position and on the room characteristics (geometry and wall properties).

On one hand, in some applications the effect of room reverberation is undesirable. For example, most of microphone array techniques and multi-loudspeaker systems are based on free field models and their performance decrease with reverberation. On the other hand, reverberation plays an important role in auditory scene synthesis, e.g. in virtual reality framework. In both cases, having all RIRs of the room could potentially be used to improve their performance or their realism.

Measuring the PAF is fundamentally a sampling problem: from a limited number of point measurements, the goal is to reconstruct (i.e. interpolate) the acoustic wavefield at any position in space and at any time.

Standard acquisition of signals relies on a regular sampling of space and time with respect to Shannon-Nyquist theory. At a given temporal frequency, the space sampling has to be dense enough to avoid aliasing in reconstruction and interpolation [1]. However, as we shall see later, such a direct measurement of a time-varying 3D image often requires an extremely high number of microphones. Nevertheless, informed by the physical nature of the measured signal, we can reduce the number of sampling locations, even for rooms with unknown geometry. This number is directly linked to the number of microphones if one wants to acquire the signals simultaneously, in a microphone array setting. In ref. [2] a method based on Dynamic Time Warping is used for the interpolation of the early part of the RIRs. Another example is given in ref. [3] that uses an acoustic model of rooms. This model is based on the modal theory and assumes that all RIRs share the same damped complex sinusoids (associated to common poles) with different amplitudes (residues). After the estimation of poles, their residues are estimated for each source position on a line considering a space dependency as a cosine function. Whereas the first method of [2] can interpolate the early part of the RIRs, the second one of [3] is adapted to the interpolation of the whole RIRs at low frequencies along a line.

In this paper we study the sampling and the interpolation of RIRs, at low frequencies, within a whole 3D domain Ω of the space, using the *Compressed Sensing* (CS) paradigm: this principle allows the reconstruction of signals from a limited number of measurements, if the signal is sparse (exactly or approximately) in some domains. In the case of room acoustics, this sparsity property is based on the modal theory. Although based on a different principle, the proposed method can be seen as an extension of ref. [3], adapted for 3D domains. It is important to note that this 3D interpolation is

Manuscript received ?? ??, ??; revised ?? ??, ??; accepted ?? ??, ???. Date of publication ?? ?? ??; date of current version nulldate. This work was supported by the Agence Nationale de la Recherche (ANR), project ECHANGE (ANR-08-EMER- 006). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Woon-Seng Gan.

R. Mignot is with the Institut Langevin (ESPCI ParisTech, CNRS UMR 7587, Paris Diderot University), 75238 Paris cedex 05, France, and also with the Institut Jean Le Rond d'Alembert (CNRS UMR 7190, Université Pierre et Marie Curie), 75252 Paris cedex 05, France, and also with the Department of Signal Processing and Acoustics (School of Electrical Engineering, Aalto University), Espoo FI-00076, Finland (e-mail: remi.mignot@aalto.fi).

G. Chardon is with the Institut Langevin (ESPCI ParisTech, CNRS UMR 7587, Paris Diderot University), 75238 Paris cedex 05, France, and also with Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, 1040 Wien Austria. The work of G. Chardon is supported by Austrian Science Fund (FWF) START-project FLAME ("Frames and Linear Operators for Acoustical Modeling and Parameter Estimation", Y 551-N13).

L. Daudet is with the Institut Langevin (ESPCI ParisTech, CNRS UMR 7587, Paris Diderot University), 75238 Paris cedex 05, France (e-mail: laurent.daudet@espci.fr). The work of L. Daudet is on a joint affiliation with the Institut Universitaire de France, and is supported by LABEX WIFI under references ANR-10-LABX-24 and ANR-10-IDEX-0001-02 PSL*.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org. Digital Object Identifier ??????

here performed without the explicit knowledge of the room shape - that should only satisfy some general assumptions detailed in sections III and IV.

The outline of this paper is as follows. Section II recalls the basics of uniform sampling, and discusses how it applies to the sampling of the 4D-FT spectrum of the PAF (3D in space, 1D in time). We observe that in the case of a 3D sampling of the space, the spectrum essentially lies on a 3D hypersurface, and not in the whole 4D volume. In section III, this observation is accounted for by theoretical results. A first sparsity property of the PAF is exhibited and a first approach is proposed to sample it within a 3D volume with few measurements. In section IV, a second sparsity model is derived, based the modal analysis of a rectangular room. This model allows the use of a *Compressed Sensing* framework for the sampling. Section V presents some details on the proposed algorithm implementations. In particular, some strategies must be designed to circumvent the large computational requirements of the algorithm. Numerical and experimental results are presented in section VI, and show the relevance of this approach in practical settings. Concluding remarks are finally presented in section VII.

The sparse model and a small subset of the results have previously been presented in a conference paper [4]. The main novelties of this paper are: a better theoretical justification of the model, more detailed explanation of the proposed algorithms, the use of cross-validation in section III-C, and a deeper analysis of the experimental data.

II. Uniform sampling

The Green's function, for the wave equation with boundary conditions, gives a complete description of the acoustic transfer between any source and receiver, within a given room. In [1], it is renamed Plenacoustic Function (PAF), and it can be described as the set of all Room Impulse Responses (RIRs), for all source / receiver positions. Note that, due to reciprocity properties, if sources and receivers are omnidirectional, they play symmetric roles.

In this section, considering a fixed source (as this corresponds to our experimental setup), we recall how standard sampling of the PAF can be done within a volume Ω of the space, using a uniform 3D microphone array, as a function of the position $\vec{X} = [x, y, z]^T$ of the receiver.

The primary design parameter is the temporal bandwidth that is required for the applications at hand. If the maximum frequency is fixed at f_c [Hz], higher frequencies are removed with an anti-aliasing low-pass filter and, assuming ideal filters, the sampling in time is done at a rate $F_s > 2f_c$. Depending on this temporal frequency bandwidth, the distance between microphones (sampling in space) has to be small enough to avoid spatial aliasing. In this section we present the spectrum of the PAF to define a criterion for the sampling. The RIRs will be denoted by the space/time dependent function: $p(t, \vec{X})$.

A. Spectrum and sampling

In [1], Ajdler studied the spectrum of the PAF on a line parallel to the (Ox) axis. With ω [rad.s^{-1}] the temporal angular frequency and φ_x [rad.m^{-1}] the spatial angular frequency, he observed that the energy of the 2D-FT $\hat{p}(\omega, \varphi_x) = \mathrm{TF}\{p(t, x)\}$ is mainly concentrated within the triangle bounded by $|\varphi_x| \leq |\omega|/\mathrm{c}_0$, which corresponds to the dispersion relation of propagative waves. Then, for growing φ_x, he demonstrated that $\hat{p}$ decreases faster than an exponential for $|\varphi_x| > |\omega|/\mathrm{c}_0$, which corresponds to evanescent waves. From this, he determines a sampling theorem, which describes how to sample in space for a target Signal-to-Noise Ratio SNR_0:

$$\frac{2\pi}{\delta_x} > \frac{2\omega_c}{c_0} + \varepsilon(\mathrm{SNR}_0, \omega_c), \tag{1}$$

where δ_x is the spatial sampling step on the line, $\omega_c = 2\pi f_c$ is the cutoff frequency, c_0 is the sound velocity. $\varepsilon(\mathrm{SNR}_0, \omega_c)$, whose exact expression is given in [1], accounts for the influence of evanescent waves. Under the far field assumption, and sufficiently far from the walls, evanescent waves are negligible, which leads to $\varepsilon = 0$. Then, in this case, the spectrum of the PAF is included within the triangle of equation $\varphi_x^2 \leq \omega^2/\,{\mathrm{c}_0}^2$, and the sampling theorem becomes: $\delta_x < \pi c_0/\omega_c$.

In the case of 2D sampling (in a plane parallel to (Oxy)), under the far field assumption, the 3D-FT of the PAF, $\hat{p}(\omega, \varphi_x, \varphi_y)$, has its support in the cone of equation $\varphi_x^2 + \varphi_y^2 \leq \omega^2/c_0^2$, where φ_x and φ_y are the spatial frequencies along axes (Ox) and (Oy). We have a similar result in the case of a 3D sampling: the support of the spectrum of the PAF $\hat{p}(\omega, \varphi_x, \varphi_y, \varphi_z)$ is essentially such that $\varphi_x^2 + \varphi_y^2 + \varphi_z^2 \leq \omega^2/c_0^2$; which means that it is included inside a hypercone.

Finally, in any case, to avoid spatial aliasing we have to choose sampling steps that satisfy the sampling theorem:

$$\delta_v < \frac{\pi c_0}{\omega_c}, \ \forall v \in \{x, y, z\}\,. \tag{2}$$

B. Reconstruction

The sampling of the PAF gives $p(t_n, \vec{X}_m)$ for $t_n = n/F_s$ and $\vec{X}_m$ on a spatial grid. The reconstruction of the RIRs for any time and position is done using a 4D interpolation filter, which may be separable in time and space.

In theory, the ideal reconstruction should be performed using convolution with a sinc function which has infinite support, therefore requiring an infinite number of sampling points, in time and space. Because of the exponential time decay of the RIRs, the responses can be truncated in time, and using finite length filters provides good approximations.

However, in space this problem remains, because a precise interpolation requires an overly large number of microphones. Actually, in order to reconstruct the RIRs within a finite sub-domain Ω of the room, in practice there are 2 possible strategies:

- by fixing the spatial sampling step δ according to Shannon-Nyquist requirements, one has to increase the order of the 3D interpolation filter (in space) in order to improve the reconstruction. Consequently, the microphone array must be larger than Ω, and according to the desired quality, the number of microphones may be unrealistic in practice.
- by fixing the size of the array, one can improve the quality by taking a finer grid. Even if the array does not become bigger, the number of microphones increases, and

boundary effects may still be present. As in the previous case, the number of microphones may become very high in practice.

Alternatively, it is also possible to make a compromise between these two strategies.

C. Sparse spectrum

As discussed above, for a 3D problem the support of the spectrum of the PAF is included inside a 4D hypercone, for which the axis of revolution is the temporal frequency axis ω.

Further observations reveal that, whereas 1D or 2D samplings of the space (on a line or a surface respectively) give a full spectrum lying inside a triangle or a cone respectively, a 3D sampling in a volume gives an almost empty spectrum, for which the energy is concentrated on the surface of the hypercone of equation $\varphi_x^2+\varphi_y^2+\varphi_z^2=\omega^2/\,{\rm c_0}^2$. Equivalently, for a given frequency ω, the spectrum is on the surface of the sphere of radius $|\omega|/\,{\rm c_0}$, which corresponds to the set of plane waves with wavenumber $|\omega|/\,{\rm c_0}$ and wavelength $2\pi\,{\rm c_0}\,/|\omega|$.

An equivalent observation, and somehow easier to visualize, can be done for a 2D problem, which represents wave propagations on an elastic membrane for example. Using synthetic signals, figure 1a presents the 2D-FT spectrum of the PAF sampled on a line. The support is a full triangle of equation $\varphi_x^2 \leq \omega^2/c_0^2$. Figures 1(b,c,d) present 3 different sections of the 3D-FT spectrum of the PAF sampled on a square surface. In this 2D problem, this quasi-empty spectrum lies only on the surface of the cone of equation $\varphi_x^2+\varphi_y^2=\omega^2/c_0^2$. Figure 1d shows that for a given frequency ω, the spectrum is on the circle of radius $|\omega|/\,{\rm c_0}$ which corresponds to the associated plane waves.

As a consequence, for a 3D problem, and sufficiently far from the source and the walls so that evanescent waves can be neglected, the 4D-FT spectrum of the PAF sampled within a spatial volume Ω does not fill a 4D volume of the 4D frequency space (ω, φ_x, φ_y, φ_z) but lies on a 3D surface, which is an hypercone. The approaches of the next section exploit this property in order to derive new sampling algorithms which need less measurements in order to reconstruct the RIRs within a volume Ω of the space.

III. Structured sparsity

A. Modal decomposition

Considering linear acoustic propagation away from the sources, the acoustic pressure $p(t,\vec{X})$ is governed by the *wave equation*

$$\Delta p(t,\vec{X}) - \frac{1}{{\rm c_0}^2}\frac{\partial^2}{\partial t^2}p(t,\vec{X}) = 0, \tag{3}$$

where $\Delta=\nabla^2$ is the Laplacian operator. At low frequencies, assuming a modal behavior for closed rooms, the solution can be decomposed as a discrete sum of damped complex harmonic signals with the angular frequencies ω_q:

$$p(t,\vec{X}) = \sum_{q\in\mathbb{Z}^\star} A_q\ \phi_q(\vec{X})\ g_q(t), \tag{4}$$

where the A_q's are complex coefficients, ϕ_q is the modal shape of mode q, and $g_q(t)$ is the corresponding time evolution, with $g_q(t) = {\rm e}^{jk_q\,{\rm c_0}\,t} = {\rm e}^{\xi_q t}\,{\rm e}^{j\omega_q t}$ for $t\geq 0$ and $g_q(t)=0$ for negative time. Finally,$k_q=(\omega_q-j\xi_q)/{\rm c_0}$ is the wavenumber of mode q with ω_q its angular frequency and $\xi_q<0$ its damping coefficient. Note that the ω_q's, ξ_q's and ϕ_q's depend on the boundary conditions (room geometry and wall properties), while the A_q's depend on the initial conditions. Because in our case the source position is not known and the source signal is an impulse at $t=0$, we only consider the homogeneous wave equation (3), without source, which implies initial conditions in $t=0$.

From (3) and (4), and using the orthogonality of the functions $g_q(t)$, we get the Helmholtz equation for every mode:

$$\Delta\phi_q + k_q^2\phi_q = 0. \tag{5}$$

Note that the orthogonality property of the functions $g_q(t)$ is fully validated when $\xi_q=0$, i.e. for ideally rigid walls. In the case of non-rigid walls, we make the usual assumption that eq. (5) remains valid at least far from the walls.

B. Plane wave approximation

In the Helmholtz equation, ϕ_q is the eigenmode of the Laplacian operator with eigenvalue $-k_q^2$. For a real k_q (rigid walls), if the room is star-shaped (note that this includes convex rooms), previous studies (cf. [5]) have shown that an eigenmode of the Laplacian with a negative eigenvalue can be approximated by a finite sum of plane waves incoming from various directions, and sharing the same wavenumber k. Then

$$\phi_q(\vec{X}) \approx \textstyle\sum_{r=1}^{R} a_{q,r}\,{\rm e}^{j\vec{k}_{q,r}\cdot\vec{X}} \tag{6}$$

is the R-order approximation of ϕ_q, with $\vec{k}_{q,r}$ the 3D wavevector r of the mode q, such that $\|\vec{k}_{q,r}\|_2=|k_q|$. More details are given in appendix A. For damping walls, in theory the losses modify ϕ_q, nevertheless we assume that the approximation (6) remains valid, at least for $\vec{X}$ far from the walls.

Consequently, considering a finite frequency range $[0,\omega_c]$ containing Q real modes, or equivalently $2Q$ complex modes, and considering R-order approximations of the ϕ_q's, the RIR $p(t,\vec{X})$ can be approximated by a sum of $2QR$ damped harmonic plane waves, $\exp(j(k_q\,{\rm c_0}\,t+\vec{k}_{q,r}\cdot\vec{X}))$, with complex wavenumber $k_q=(\omega_q-j\xi_q)/{\rm c_0}$ and real wavevectors $\vec{k}_{q,r}$ such that $\|\vec{k}_{q,r}\|_2=|\omega_q|/\,{\rm c_0}$, and with coefficients linked by the relation $\alpha_{q,r}=A_q a_{q,r}$.

Note that this approximation, coming from [5], theoretically validates the observation of the sparse 4D-FT spectrum of the PAF given in section II-C. We speak about *Structured Sparsity*, first because of the modal representation of eq. (4), which is sparse in the time domain, and second because the 4D-FT spectrum is concentrated on a 3D surface:for every frequency ω_q, the modal shape ϕ_q is modeled by a sum of plane waves for which the wavevectors lie on the sphere of radius $|\omega_q|/\,{\rm c_0}$ only (i.e. $\hat{p}(\omega,\vec{k})\approx 0$ for $\|\vec{k}\|_2\neq|\omega|/\,{\rm c_0}$). But note that it does not assume the sparsity of ϕ_q in a dictionnary of plane waves: the finite sum of eq. (6) is an approximation required for computation.

In [3], the modal shape is assumed to be a cosine function on a line (Ox), with wavenumber $k_x\leq|k_q|$, which corresponds to the sum of two equivalent plane waves. In this section, the

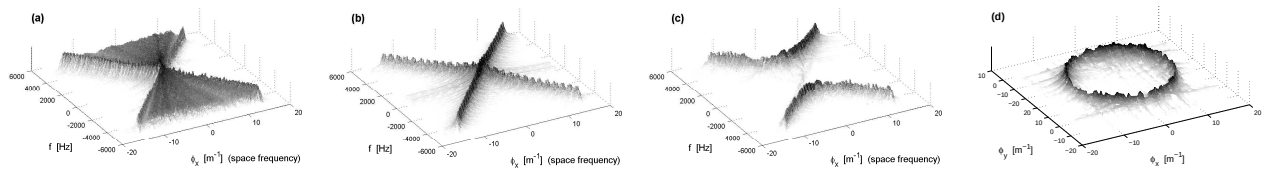


Fig. 1. Spectrum of the PAF for a 2D problem (synthetic signals). (a): 2D-FT of the PAF sampled on a line. (b,c,d): 3D-FT of the PAF sampled on a square; (b): $\varphi_y = 0$ m^{-1}, (c): $\varphi_y = 7$ m^{-1}, (d): $f = 4$ kHz. As expected for a 2D problem, the sampling of the PAF on a line gives a full spectrum, and its sampling on a 2D domain gives a sparse spectrum which lies on a cone, for which the symmetry axis is the temporal frequency axis f.

modal shape in 3D is represented, in a more general way, as a sum of R plane waves, which is valid with a weaker assumption (star-shaped room).

C. Algorithm

Now, taking advantage of this Structured Sparsity, we present an algorithm previously proposed for the interpolation of impulse responses of plates [6]. First, using an array of M microphones placed at the $\vec{X}_m$ sampling points within Ω (with uniform or random sampling, cf. sec. IV-B), we acquire the digital signals $p(t_n, \vec{X}_m)$, of length N samples each. Second, we can reconstruct the RIRs in Ω using the following algorithm:

(a) The shared wavenumbers k_q are estimated using a joint estimation of damped sinusoidal components (using for example the algorithms MUSIC [7], ESPRIT [8], or SOMP [9]). Note that this stage corresponds to a sparse decomposition of the M signals using a joint sparsity model along the temporal frequencies, with a dictionary of damped sinusoids, or equivalently of common acoustic poles $p_q = j\,\mathrm{c}_0\,k_q$ in the Laplace domain.

(b) From (4), the $(M \times N)$ matrix S of signals, such that $S_{[m,n]} = p(t_n, \vec{X}_m)$, can be written as $S = \Phi G$, where Φ is the $(M \times Q)$ matrix of modes, $\Phi_{[m,q]} = A_q \phi_q(\vec{X}_m)$, and G is the $(Q \times N)$ dictionary of damped exponentials, $G_{[q,n]} = \mathrm{e}^{j k_q \mathrm{c}_0 t_n}$. Then, with $Q < N$, the modal shape matrix Φ is estimated using the ℓ_2 optimization:

$$\tilde{\Phi} = P G^H (G G^H)^{-1}. \tag{7}$$

(c) Now from (6), $\phi_q \approx \psi_q \alpha_q$ where ψ_q is the $(M \times R)$ matrix of the plane waves, $\psi_{q\,[m,r]} = \mathrm{e}^{j \vec{k}_{q,r} \cdot \vec{X}_m}$, sharing the same wavenumber k_q. The $\vec{k}_{q,r}$'s are chosen using a uniform sampling of the sphere of radius $|\omega_q|/\mathrm{c}_0$, cf. [10], [11]. Then, with $M > R$, the coefficients $\alpha_{q,r}$ of the vector α_q are estimated using the least squares projection of every $\tilde{\phi}_q$ into the corresponding basis of ψ_q as follows:

$$\tilde{\alpha}_q = (\psi_q^H \psi_q)^{-1} \psi_q^H \tilde{\phi}_q. \tag{8}$$

(d) Finally, the RIRs can be interpolated for any $t \in [0, N/F_s]$ and any position $\vec{X} \in \Omega$ using the approximation:

$$\tilde{p}(t, \vec{X}) = \sum_{q,r} \tilde{\alpha}_{q,r}\ \mathrm{e}^{j(k_q \mathrm{c}_0 t + \vec{k}_{q,r} \cdot \vec{X})}. \tag{9}$$

Note that S is a real matrix, hence the coefficients of the complex modes have to obey the Hermitian symmetry. This implies: $\alpha_{q,r} = \alpha^*_{-q,r}$, $k_q = -k^*_{-q}$ and $\vec{k}_{q,r} = -\vec{k}_{-q,r}$, where the symbol $.^*$ denotes the conjugate. In practice, this Hermitian symmetry is used in stages (b) and (c) in order to reduce the size of matrices.

In stage (c), the sphere of radius $|\omega|/\mathrm{c}_0$ is sampled using R plane waves. The choice of R is important, because a small R produces bad approximations, but a value that is too high (still respecting $R < M$) leads to overfitting: it gives good approximations for the measured positions of microphones, but interpolations with poor quality. In section VI, two methods (CS0 and CS1) are tested and compared:

- **CS0**: First, we have empirically determined $R \approx 3M/4$ for all modes. This value gives a good conditioning for the computation of the pseudo-inverse of ψ_q, cf. (8), and some informal tests reveal that it gives good results in most of the cases.
- **CS1**: Second, we modified the previous algorithm to choose the best R for every mode q separately, using a cross-validation procedure. A small number m of microphones are randomly selected among the M microphones of the array. Then for every value of $R < M - m$, the vector $\tilde{\alpha}_q$ of (8) is computed using the $M - m$ other microphone positions. Finally the modal shape ϕ_q are reconstructed at the m selected positions, and the value of R which gives the lowest error is chosen.

Note that the choice of the number Q of estimated modes is not a critical issue. In practice, we can use an approximate value of the room volume V and the relation $Q \approx 4\pi V (f_c/\mathrm{c}_0)^3/3$ (cf. [12]). Nevertheless, if $Q > N$, stages (a) and (b) cannot be performed, as only the time information is exploited in these stages. The next section presents a stronger sparsity property, which takes into account simultaneously the information of time and space, but with a restricted assumption on the room geometry.

IV. Plane wave sparsity

In this section, we study the solutions of the wave equation in the simple case of a rectangular room. From this study, we exhibit a stronger property of sparsity which justifies the use of the Compressed Sensing framework (CS). The derived algorithm, detailed in sec. V, will be named **CS2**.

A. Modal analysis in a rectangular room

In the case of a rectangular room with rigid walls, we can make the variable separation in Cartesian coordinates (x, y, z), cf. e.g. [12]. Then, each modal shape is written as the product

of 3 functions of one variable. With $\vec{X} = [x, y, z]^T$, the RIRs become:

$$p(t, \vec{X}) = \sum_{q \in \mathbb{Z}^\star} A_q \, F_{xq}(x) \, F_{yq}(y) \, F_{zq}(z) \; \mathrm{e}^{j k_q \, \mathrm{c}_0 \, t} \, . \qquad (10)$$

For each mode q, these functions verify the 1D Helmholtz equation $\partial_v^2 F_v + k_v^2 F_v = 0$ for $v \in \{x, y, z\}$. With rigid walls, the k_v's are real constants such that $k_x^2 + k_y^2 + k_z^2 = k^2$ (cf. [12]). According to the Helmholtz equation, for each Cartesian coordinate v the F_v's are the sum of 2 solutions: $F_v(v) = A_v^+ \, \mathrm{e}^{j k_v v} + A_v^- \, \mathrm{e}^{-j k_v v}$. Then, expanding $F_x F_y F_z$, the modal shape $\phi_q(\vec{X})$ is written as the sum of 8 plane waves $\mathrm{e}^{\pm j k_x x \pm j k_y y \pm j k_z z} = \mathrm{e}^{j \vec{k} \cdot \vec{X}}$, with $\vec{k} = [\pm k_x, \pm k_y, \pm k_z]^T$.

In the case of non-rigid walls, as the wavenumber k is complex: $k = (\omega - j\xi)/\mathrm{c}_0$, the k_v's are complex too. This implies a slight decrease of the F_v's near the walls. Nevertheless, for $\vec{X}$ far from the walls, we assume that the imaginary part of k_v is negligible, and that $k_x^2 + k_y^2 + k_z^2 = \mathrm{Re}(k)^2 = \omega^2 / {\mathrm{c}_0}^2$.

Note that in the case of a rectangular room, the wavevectors $\vec{k} = [\pm k_x, \pm k_y, \pm k_z]^T$ are at the vertices of an inscribed parallelepiped of the sphere with radius $|\omega|/\mathrm{c}_0$. Moreover, whereas the modal density, which is related to the number of modes per frequency range, strongly increases with the frequency, with f^2, all the wavevectors are uniformly spaced in the $\vec{k}$-space (with coordinates k_x, k_y, k_z), cf. [12].

Consequently, in a bandwidth containing $2Q$ complex modes, the RIRs can be written as the sum of $16Q$ harmonic plane waves in the case of rectangular rooms. Note that in the previous section, each modal shape was approximated by R fixed plane waves sampling uniformly the sphere of radius $|\omega|/\mathrm{c}_0$, whereas here, with the assumption of rectangular room, only 8 plane waves are required by mode. Then, this stronger sparsity property justifies the use of CS techniques.

This model is, strictly speaking, only valid for rectangular rooms. In spite of this strong restriction compared to the weaker assumption of the methods CS0 and CS1 (star-shaped rooms), this model remains interesting because rectangular rooms are very often used, and as shall be seen in the experimental section VI, it gives better results than CS0 or CS1 in hard conditions (low signal-to-noise, interpolation on a larger bandwidth).

B. Compressed Sensing framework

The general problem consists in the reconstruction of a signal $y \in \mathbb{R}^{\mathcal{N}}$ from $\mathcal{M}$ observations x_m, linked by the linear system $x = \Phi y$. *Compressed Sensing* (CS) deals with the underdetermined case, for which there are more unknowns than equations ($\mathcal{N} > \mathcal{M}$), cf. e.g. [13], [14]. As such a problem cannot be solved without additional hypothesis, the underlying idea is that if y lives in a subspace of dimension $\mathcal{K}$ and with basis ψ, for $\mathcal{K} < \mathcal{M}$, we can solve $y = \psi a$ writing $x = \Phi y = \Phi \psi a = \theta a$. However, in general we do not know ψ.

Then, we define $\mathcal{L}$ vectors ψ_l, forming the matrix Ψ with $\mathcal{L} \gg \mathcal{K}$, and we look for a basis which explains y. In other words, we look for a vector $\alpha \in \mathbb{R}^L$ K-sparse (where no more than K coefficients are non-zero), such that $y = \Psi \alpha$. Unfortunately, the problem of finding the sparsest solution is not convex, and hence difficult to solve. However, we can change it into a convex problem by considering the following *Basis Pursuit Denoising* approach:

$$\min_{\alpha \in \mathbb{R}^{\mathcal{L}}} \|\alpha\|_{\ell_1} \qquad \text{subject to} \qquad \|x - \Phi\Psi\alpha\|_{\ell_2} \leq \varepsilon, \qquad (11)$$

where the norm ℓ_n is given by $\|y\|_{\ell_n} = (\sum_i |y_i|^n)^{1/n}$, and ε is a data fidelity parameter. A high ε allows a stronger sparsity of α, and a small ε improves the reconstruction of y.

Some theoretical results (cf. e.g. [15], [16], [17]) give a sufficient condition for reconstructing y in the case of sparse signals, by the so-called Restricted Isometry Property (RIP). It quantifies how Φ and Ψ are mutually incoherent with respect to their use on sparse signals. In practice, the RIP is difficult to check, but it is verified with high probability for some random sampling matrices. In practice, this encourages the use of randomly selected observation points, which are here the microphone positions in the 3D space. Note that, conversely, a regular sampling grid might lead to a strong correlation with plane waves, whenever the wavevector gets close to be aligned to one of the x, y or z axis: such standard sampling scheme is therefore likely to be suboptimal in the CS framework.

C. Reformulation of the problem in a Compressed Sensing framework

Now, we can reformulate our problem as follows: let us define $\mathcal{S}_x$ the signal vector of the measurements $p(t_n, \vec{X}_m)$, and $\mathcal{S}_y$ the signal vector that we wish to reconstruct (interpolate) on a uniform grid of the space:

$$\mathcal{S}_{x\ [(n+1)+(m-1)N]} = p(t_n, \vec{X}_m), \quad \text{and} \qquad (12)$$

$$\mathcal{S}_{y\ [(n+1)+(s-1)N]} = p(t_n, \vec{Y}_s), \qquad (13)$$

where the $\vec{X}_m$'s are the positions of the M microphones of the array, and the $\vec{Y}_s$'s are the positions of the 3D grid. Considering the ideal reconstruction using a finely sampled uniform array (cf. sec. II), $\mathcal{S}_x$ and $\mathcal{S}_y$ are linked by $\mathcal{S}_x = \Phi_{xy} \mathcal{S}_y$, where Φ_{xy} is an interpolation matrix representing the spatial convolution for interpolating the RIRs at $\vec{X}_m$ starting from the signals on the grid of the $\vec{Y}_s$'s.

Since the number of microphones is limited in practice, we cannot directly reconstruct $\mathcal{S}_y$ from $\mathcal{S}_x$. However, thanks to the sparsity property of the RIRs as described in sec. IV-A, it is possible to solve this problem using CS. The rough idea is to define an oversized dictionary Ψ_y with harmonic plane waves which are "virtually" sampled on the grid. Then, writing $\mathcal{S}_y = \Psi_y \alpha$, in principle the problem might be solved with $\mathcal{S}_x = \Phi_{xy} \Psi_y \alpha$. Unfortunately because of the space dimensionality (4D), standard ℓ_1 optimization algorithms of (11) would require too much memory and cannot be run on standard computers. Hence, in the next section, we propose a greedy algorithm for the interpolation of the RIRs.

V. Algorithmic details

When ℓ_1 optimization procedures cannot be processed because of computational issues, greedy algorithms such as Matching Pursuit are common alternative. However, with the

size of data in this work, even this simple algorithm is too cumbersome to be computed in practice. In this section, first standard Matching Pursuit is presented, then we propose a derived version which can be applied for the sampling of the RIRs in 3D. This new algorithm is named **CS2** in this paper.

A. Matching Pursuit

Matching Pursuit (MP) [18] consists in iteratively subtracting from the signal the atom that best approximates it. This atom g is chosen among the columns of a dictionary matrix Ψ, of size $(\mathcal{M}\times\mathcal{L})$. Then the process is iterated on the residual which is, at the iteration $i+1$:

$$r_{i+1} = r_i - \alpha_i\ g_i, \tag{14}$$

with r_0 the signal to approximate, and where the vector g_i and the coefficient α_i are chosen to minimize $\|r_{i+1}\|_{\ell_2}$. If the column vectors of Ψ are normalized, the optimal atom is $g_i = \arg\max_{g\in\Psi} |\langle g, r_i\rangle|$ and the optimal coefficient is given by the correlation $\alpha_i = \langle g_i, r_i\rangle := g_i^H r_i$. The symbol $.^H$ denotes the conjugate transpose of a complex matrix or a vector.

A similar method consists in searching at each iteration a group of P atoms simultaneously minimizing the norm of the residual $r_{i+1} = r_i - G\alpha$, where G is a $(\mathcal{M}\times P)$ matrix of P atoms, and α is a $(P\times 1)$ vector. If the atoms are normalized, and if $\mathrm{rank}(G) = P$ with $P < \mathcal{M}$, the optimal matrix G_i minimizes

$$\|r_{i+1}\|_{\ell_2}^2 = \|r_i\|_{\ell_2}^2 - r_i^H G \left(G^H G\right)^{-1} G^H r_i, \tag{15}$$

and the weight vector is then $\alpha = (G^H G)^{-1} G^H r_i = G^\dagger r_i$, where the symbol $.^\dagger$ denotes the pseudo-inverse of a matrix.

In the present work, we first consider the application of Matching Pursuit considering groups of P harmonic plane waves which share the same wavenumber. For example, the rectangular room considered in the experiments, section IV-A, led to $P = 8$. Unfortunately, because of the dimensionality of the problem, it is not possible to use this algorithm as such. Indeed, among a high number of possible wavenumbers $k = (\omega - j\xi)/\mathrm{c}_0$ (that belong to a subspace of dimension 2), we would have to test a wider number of possible combinations of P plane waves on the sphere of radius ω/c_0 (in a subspace of dimension $2P$). Consequently, the matrices G live in a subspace of dimension $2 + 2P$, and exhaustive search for the most correlated group is in practice absolutely impossible. In the next section, we propose a modified algorithm which alleviates this problem.

B. Modified MP algorithm

Let us define S the $(N \times M)$ signal matrix such that $S_{[n,m]} = p(t_n, \vec{X}_m)$, and $\mathcal{S}$ its vectorized version as in equation (12), $\mathcal{S} = \mathcal{S}_x$. The residual vectors will be noted $\mathcal{R}_i$, and their $(N \times M)$ matrix versions R_i.

1) Analysis: The principle of the proposed algorithm is as follows. At every iteration i, first we choose the damped complex exponential which best approximates the M columns of R_i (which are time signal vectors), and so a wavenumber $k_i = (\omega_i - j\xi_i)/\mathrm{c}_0$ is estimated. Then, among W selected atoms, we choose a group of P harmonic plane waves (on the sphere of radius ω_i/c_0) which efficiently explains the residual $\mathcal{R}_i$, with $P < W$. For more details, the 4 stages of the iteration i are detailed here:

(A) This stage is similar to the search of poles of SOMP [9]. We define the $(N \times L_t)$ time dictionary matrix Θ with L_t columns θ_ℓ which are damped complex exponentials: $\Theta_{[n,\ell]} = \theta_{\ell[n]} = \mathrm{e}^{\xi_\ell t_n}\,\mathrm{e}^{j\omega_\ell t_n}$, with $0 < \omega_\ell \leq \omega_c$ and $\xi_\ell < 0$. Then defining the $(L_t\times M)$ correlation matrix $\eta_i := |\bar{\Theta}^H \mathrm{R}_i|$, we choose the index ℓ_i which maximizes the sum of energies: $\sum_{m=1}^{M} (\eta_{i[\ell,m]})^2$. Here, the matrix $\bar{\Theta}$ corresponds to Θ where the columns are individually normalized: $\bar{\theta}_\ell = \theta_\ell / \|\theta_\ell\|_{\ell_2}$.

(B) With the estimated wavenumber k_{ℓ_i}, we define an $(MN\times L_s)$ dictionary matrix Δ_i with L_s columns $\delta_{i,\ell}$ which are harmonic plane waves: $\delta_{i,\ell\ [(n+1)+(m-1)N]} = \mathrm{e}^{\xi_{\ell_i} t_n}\,\mathrm{e}^{j\omega_{\ell_i} t_n}\,\mathrm{e}^{j\vec{k}_\ell\cdot\vec{X}_m}$, with $\|\vec{k}_\ell\|_2 = \omega_{\ell_i}/\mathrm{c}_0$, $\forall \ell \in [1, L_s]$. Then, with $L_s \gg W > P$, we isolate W atoms $\delta_{i,\ell}$ which are the maxima of $\rho_{i[\ell]} := |\langle \bar{\delta}_{i,\ell}, \mathcal{R}_i\rangle|$. Note that ρ_i can be written $\rho_i = |\bar{\Delta}_i^H \mathcal{R}_i|$. Actually, because of possible lobes, ℓ must index a 2D grid of the uniformly sampled sphere of radius $\omega_{\ell_i}/\mathrm{c}_0$, and the chosen atoms are the W higher local maxima.

(C) Among these W atoms, we test all combinations of P atoms (there are $\binom{W}{P}$ possible combinations). Then we choose the combination G_i which minimizes (15): $\|\mathcal{R}_{i+1}\|_{\ell_2}^2 = \|\mathcal{R}_i\|_{\ell_2}^2 - \mathcal{R}_i^H \bar{G}\bar{G}^\dagger \mathcal{R}_i$, with G an $(MN\times P)$ matrix of one combination of P vectors.

(D) Finally, the best combination G_i is subtracted. Actually, here we have to consider the hermitian symmetry for real signals, and so defining $\mathcal{G}_i = [G_i, G_i^*]$, the residual $i+1$ is: $\mathcal{R}_{i+1} = \mathcal{R}_i - \bar{\mathcal{G}}_i \alpha_i$, with $\alpha_i = \bar{\mathcal{G}}_i^\dagger \mathcal{R}_i$,

Compared to the standard Matching Pursuit algorithm presented in section V-A, this new algorithm has a tremendously reduced complexity: whereas the dimension of the dictionary matrix Ψ of standard MP is $(MN \times L_s L_t)$, thanks to the variable seperation, the modified algorithm uses the matrices Θ and Δ_i with reduced dimensions $(N\times L_t)$ and $(MN\times L_s)$ respectively. Moreover, in the second stage, the atoms are individually tested on a sphere (subspace of dimension 2), which facilitates the process; and only the W best atoms are selected for the stage (C). In practice, W is chosen such that the number $\binom{W}{P}$ of possible combinations remains reasonable. With $P = 8$, a typical choice is $W = 16$, which leads to 12,870 combinations. With the standard MP algorithm, the number of combinations to test is $L_t\binom{L_s}{P}$.

As in sec. III-C, the number of modes Q is estimated beforhand using an approximate measurement of the room volume. This number Q determines the number of iterations.

Note that a slight improvement has been done by adding a additional stage. Starting from the group of P plane waves selected at the end of stage (C), the wavenumber and the

wavevectors are refined using a non-linear iterative optimization, based on a simplex search method [19].

2) Projection and interpolation: At the end of the Q iterations, we get $V = QP$ estimated harmonic plane waves, with wavenumber k_v and wavevector $\vec{k}_v$. They define the atoms of the $(MN \times V)$ basis matrix B_x:

$$\begin{aligned} B_{x\ [(n+1)+(m-1)N,\ v]} &= \mathrm{e}^{jk_v t_n}\, \mathrm{e}^{j\vec{k}_v \cdot \vec{X}_m}, \\ &\text{with } \|\vec{k}_v\|_2 = \omega_v / \mathrm{c}_0 > 0, \\ &\text{and } k_v = (\omega_v - j\xi_v)/\mathrm{c}_0\,. \end{aligned} \tag{16}$$

Then with $A_x := [B_x, B_x^*]$, considering positive and negative frequencies, we could solve the optimal solution $\tilde{\mathcal{S}}_x = A_x a$ in the mean least squares sense with $a = A_x^\dagger \mathcal{S}_x$, which would require complex calculus. In order to reduce memory requirements and makes computation faster, it is preferable to manipulate only real coefficients. Then a is obtained as follows:

$$\begin{aligned} a_{[v]} &= a^*_{[v+V]} = \mu_{[v]} + j\mu_{[v+V]}, \\ &\text{with } v \in [1, V] \\ &\text{and } \mu = \frac{1}{2}\Big[\, \mathrm{Re}\{B_x\}, -\mathrm{Im}\{B_x\} \,\Big]^\dagger \mathcal{S}_x. \end{aligned} \tag{17}$$

If the problem of (17) is ill-conditioned, in practice we remove some atoms of B_x which are linearly close to some others. For that, selecting the indexes (v_1, v_2) of the maximum of the matrix $C - I_V$, where I_V is the identity and $C := \bar{B}_x^H \bar{B}_x$ is the normalized correlation matrix, we remove the plane wave $v \in \{v_1, v_2\}$ which minimizes $\rho_v = |\langle b_v, \mathcal{S} \rangle|$. This process is iterated until the problem gets well-conditioned. Note that the use of an orthogonal projection in stage (D) (as with the *Orthogonal Matching Pursuit [18]* algorithm) would partly solve this issue, but the associated computational cost would be prohibitive for the problem at hand.

Finally, the interpolation at any position $\vec{Y} \in \Omega$ and any time $t \in [0, N/F_s]$, is done by:

$$\tilde{p}(t, \vec{Y}) = \sum_{v=1}^{2V} a_{[v]}\ \mathrm{e}^{jk_v t}\, \mathrm{e}^{j\vec{k}_v \cdot \vec{Y}} \tag{18}$$

or $\tilde{\mathcal{S}}_y = A_y a$ where A_y corresponds to the matrix basis of harmonic plane waves at the position $\vec{Y}$. Note that while the matrices are normalized in section V-B1, B_x and A_y are not normalized in (17) and (18).

C. Improvements for fast computation

Taking advantage of the variable separation (in time and space), we can significantly reduce the matrix dimensions and the number of floating-point operations which are respectively associated to the used memory size and CPU usage. The following points allow the computation of the algorithm with a reasonable time and memory size:

- In stage (A), and equivalently in SOMP, if the frequency axis of ω is uniformly sampled between 0 and $F_s/2$, instead of handling the matrix Θ and computing the product $\bar{\Theta}^H \mathrm{R}_i$, we compute the Fast Fourier Transforms in time of: $\mathrm{e}^{\xi t_n}\, \mathrm{R}_i[n, m]$, alternatively for every sampled damping coefficients ξ.
- For the computation of ρ_i in stage (B), we prove that
$$|\bar{\Delta}_i^H \mathcal{R}_i| = |\bar{\theta}_{\ell_i}^H\, \mathrm{R}_i\, \bar{\Sigma}_i^*|^T, \tag{19}$$
where θ_{ℓ_i} is the $(N \times 1)$ vector chosen in stage (A), and Σ_i is an $(M \times L_s)$ space dictionary matrix such that $\Sigma_{i[m,\ell]} = \mathrm{e}^{j\vec{k}_\ell \cdot \vec{X}_m}$, with $\|\vec{k}_\ell\|_2 = \omega_{\ell_i} / \mathrm{c}_0$. Whereas the first member requires at least NML_s floating-point numbers, the second one requires significantly less memory, $N + ML_s$ numbers. Moreover, the time of computation is significantly reduced because the construction of $\bar{\theta}_{\ell_i}$ and $\bar{\Sigma}_i$ is faster than this one of $\bar{\Delta}_i$.
- In stage (C), we have to select the group of plane waves which minimizes $\|\mathcal{R}_{i+1}\|^2_{\ell_2}$. Proving that
$$\mathcal{R}_i^H \bar{G}_i \bar{G}_i^\dagger \mathcal{R}_i = (\bar{\theta}_{\ell_i}^H \mathrm{R}_i)^* \bar{\Sigma}_i \bar{\Sigma}_i^\dagger (\bar{\theta}_{\ell_i}^H \mathrm{R}_i)^T, \tag{20}$$
the number of required floating-point operations is reduced with a factor N. Here, Σ_i is a $(M \times P)$ matrix of a candidate group of P plane waves.
- During the reconstruction of the RIRs, instead of using the formula $\mathcal{S}_y = A_y a$, we can accelerate the computation by reducing the matrix dimensions and we can simplify the reconstruction in the case of I interpolation positions $\vec{Y}_r$:
$$S_y = 2\,\mathrm{Re}\big\{ \Theta_y\, \mathrm{diag}(b) \Sigma_y^T \big\}\,, \tag{21}$$
where b is the first half of a, $\Theta_{y[n,v]} = \mathrm{e}^{jk_v t_n}$ and $\Sigma_{y[r,v]} = \mathrm{e}^{j\vec{k}_v \cdot \vec{Y}_r}$, for $r \in [1, I]$ and $v \in [1, V]$.
- Finally, for the computation of the correlation matrix C, we prove that
$$\bar{B}_x^H \bar{B}_x = (\bar{\Theta}_x^H \bar{\Theta}_x) \dot{\times} (\bar{\Sigma}_x^H \bar{\Sigma}_x), \tag{22}$$
where $\dot{\times}$ symbolises the array multiply. Whereas the first member needs VMN floating-point numbers and V^2MN operations, the second one needs only $V(M+N)$ numbers and $V^2(M+N+1)$ operations.

VI. Experiments and results

In the following, we present some results of the three algorithms presented in this paper. They will be named method CS0, method CS1 (cf. sec. III-C), and method CS2 (cf. sec. V-B). The quality of the interpolation is evaluated using the Signal-to-Noise Ratio (SNR) [dB] and the normalized Pearson correlation coefficient c [%]. With s the $(N \times 1)$ vector of the target RIR, such that $s_{[n]} = p(t_n, \vec{X})$, and $\widetilde{s}$ its interpolation:

$$\mathrm{SNR}_{dB} = 20 \log_{10} \left(\frac{\|s\|_{\ell_2}}{\|s - \widetilde{s}\|_{\ell_2}} \right), \tag{23}$$

$$C_{\%} = 100\, \frac{|\langle s, \widetilde{s} \rangle|}{\|s\|_{\ell_2}\ \|\widetilde{s}\|_{\ell_2}}. \tag{24}$$

Note that during the first stage of CS0 and CS1, we use the algorithm SOMP [9] because some preliminary tests revealed that it gives better results than the other damped sinusoidal component analysis methods. For the time dictionary Θ of SOMP and CS2, we use a grid of the possible wavenumbers $k = (\omega - j\xi)/\mathrm{c}_0$. The frequency axis of ω is uniformly sampled on $[0, \pi F_s]$, and in the experiments of this section,

Methods	M (mic. number)	SNR [dB]
uniform sampling	64	12.7
	125	19
	216	23.8
method CS0	64	15.6
	96	25.2
	125	29.2
method CS1	64	11.8
	96	17
	125	22.1
method CS2	64	14.2
	96	16.6
	125	17.8

TABLE I
NUMERICAL EXPERIMENT: COMPARISON BETWEEN UNIFORM SAMPLING AND METHODS CS0, CS1 AND CS2, ON SYNTHETIC RIRS.

the range of the damping coefficients ξ is uniformly sampled on the range $[0.5\,\xi^\star, 2\xi^\star]$, where $\xi^\star = -3\ln 10/RT_{60}$ is the damping associated to the estimated reverberation time at 60 dB, RT_{60}. Typically, we use 8192 values of ω, 128 values of ξ, and the number of plane waves used in stage (B) of sec. V-B1 is $L_s = 5000$.

This section presents successively some results on numerical simulations and measured RIRs. The estimated RT_{60} of the measures is approximately 1.25 seconds. For the numerical simulations, the reflexion coefficients of the wall have been set in order to get approximately the same reverberation time.

A. Preliminary numerical results

Table I compares the uniform sampling to the proposed methods. Here, we aim at reconstructing the RIRs within a cube Ω of side 1.7m, starting from M simulated RIRs (cf. [20], [21]) of a virtual array (regular for the uniform sampling, random for CS0, CS1 and CS2), in a rectangular room of sides (3.8, 8.15, 3.3)m. The source position is $\vec{X}_s = (3.3, 7.7, 1)$, and the center of the array is $\vec{X}_a = (1.8, 2.8, 1.6)$. The simulated RIRs are filtered by a low-pass filter of cut-off frequency $f_c = 300$Hz, cf. [21], the sampling rate is $F_s = 750$Hz, and the SNRs are averaged over 2744 interpolation positions in Ω.

Concerning the regular array which is cubic, the choosen spatial sampling step δ is the same for all directions (Ox, Oy and Oz) and it respects the sampling theorem: $\delta < \mathrm{c}_0/(2f_c)$. For a given number M of microphones (64, 125 or 216, cf. table I), the step δ and the size of the array have been chosen as follows: we tested some configurations of regular arrays corresponding to the first strategy of sec. II-B, to the second one, or to a compromise of both. The displayed result uses the configuration which provides the better performance.

We observe that, for a given number of microphones, method CS0 significantly outperforms uniform sampling (cf. $M = 64$ or 125). Equivalently, methods CS0 and CS1 can obtain equivalent performance as regular sampling, but with a smaller number of microphones (see for instance $M = 96$ for CS0, $M = 125$ for CS1, and $M = 216$ for the uniform sampling). According to these preliminary results, method CS2 does not seem to be competitive; we will see its benefits at a later stage.

B. Experimental results

We have then designed a real 3D array with 120 electret microphones, randomly positioned within a cube of size 2m (cf. fig. 2). The room has dimensions (3.9, 8.15, 3.35)m, it was empty but still had features that made it non-ideal: a doorway, two windows, a cornice, concrete walls, wood panels, etc. The source is a baffled loudspeaker placed far from the array, at $\vec{X}_s = (1.8, 7.5, 1.6)$, and the center of the array is at $\vec{X}_a = (1.9, 3.1, 1.5)$. Note that with this configuration, the sides of the array are at 50cm from the floor and 90cm from the closest wall. The RIRs have been measured using sine sweeps [22] in the bandwidth [50, 1000]Hz. The sine sweeps were long enough in order to reduce the noise of measurements. In order to isolate the modes below a cutoff frequency f_c, we have used a low-pass filter, and a downsampling at $F_s > 2f_c$.

The microphones are placed at random positions within Ω, with a statistical distribution close to uniform, up to mechanical constraints. 15 long bars are fixed (cf. fig. 2), and the microphones are at the ends of small perpendicular rods which are attached on the bars (8 per bar). The degrees of freedom are the orientations and the positions of the rods on the bars. Using synthetic RIRs, we have numerically tested a number of array configurations, respecting these mechanical constraints, and we have selected this one who produced best results. The set of microphone positions has been finely calibrated using an acoustic optimization procedure [23], with the measured positions as initial estimates.

In figure 3, three interpolated RIRs are displayed (for the three methods CS0, CS1 and CS2). One microphone of the array has been isolated for the test of the interpolation, and the analysis has been done using the 119 others. Here $f_c = 300$Hz and $F_s = 750$Hz. Figure 4 illustrates one result in the frequency domain, for method CS1. Both SNR and correlation performance measures show that the interpolated RIR is close to the measured one.

C. Parameter analysis

In figure 5, the performances are evaluated according to the number of microphones for the analysis. For the interpolation and the evaluation, we have randomly selected 15 microphones close to the center of the array (distance smaller than 80cm). The analyses have been computed with M microphones randomly chosen among the remaining positions. As a general trend, performance decreases with M. However, whereas method CS2 is almost 5dB below CS0 and CS1 at $M = 105$ microphones, the three methods are roughly equivalent for $28 \leq M \leq 40$. We also can remark that CS2 totally fails for $M < 19$. This shows that with only 46 microphones, we can reconstruct the RIRs within the 3D volume with an SNR of 15dB. The crossover between the methods is also interesting: method CS2 is a simpler model based on stronger assumptions, it is better when few information is available, on the other hand methods CS0 and CS1 have more parameters to estimate, and can therefore better explain the RIRs when a sufficiently large number of microphones is used.

Figure 6 shows the performance of the interpolation according to the distance between the interpolation position and the

Fig. 2. Pictures of the experimental microphone array. The 120 electret microphones are at the ends of the small rods, randomly placed and oriented on 15 fixed bars. The microphones are omnidirectional in the used bandwidth.

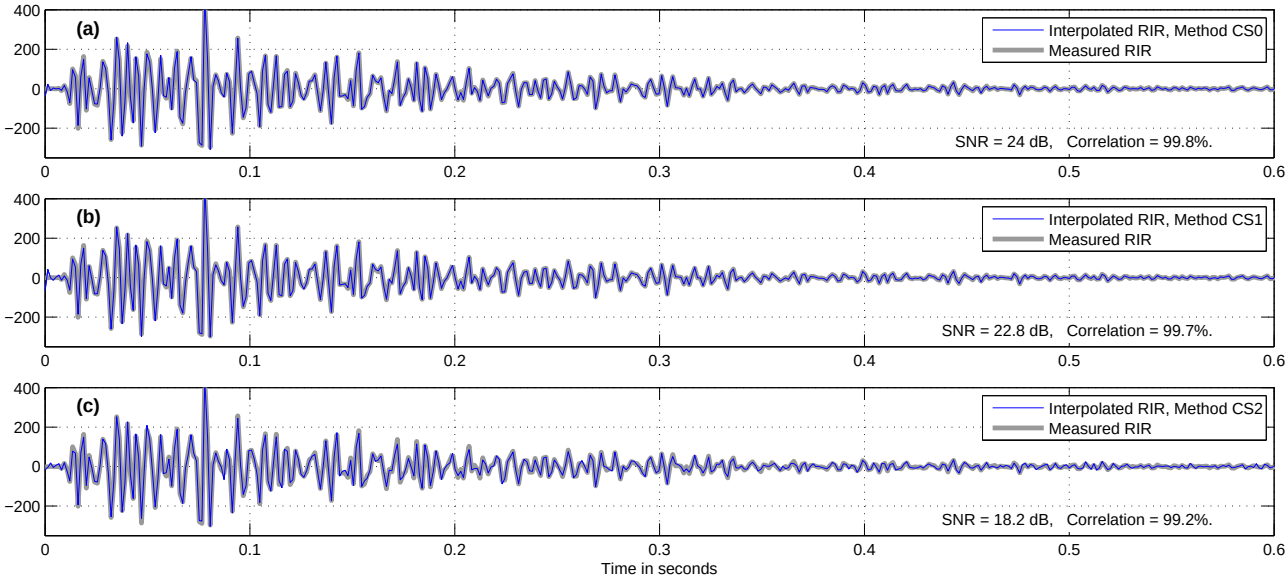


Fig. 3. Measured and interpolated RIRs. (a): Method CS0, (b): Method CS1, (c): Method CS2. On measured RIRs.

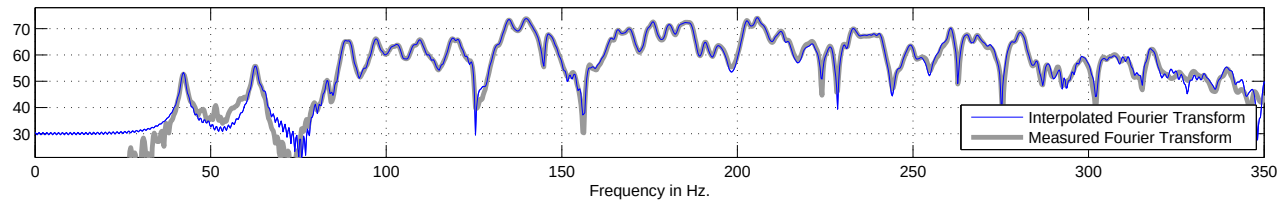


Fig. 4. Fourier transform of RIRs for the method CS1 (cf. fig. 3b). The y-axis is in a dB scale. On measured RIRs.

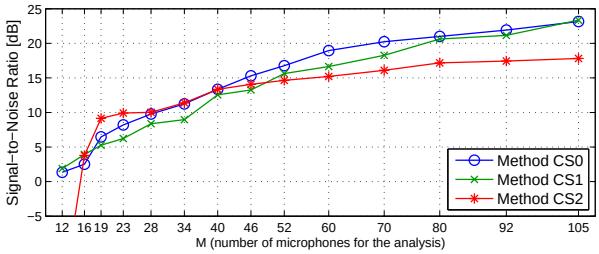


Fig. 5. Interpolation performance as a function of the number of microphones of the array. On measured RIRs.

center of the array. Here, each measured RIR is interpolated using the model parameters from the analysis of the remaining 119 measured signals. We used $f_c = 300$Hz, and $F_s = 750$Hz. The microphones are grouped according to their distance from the center of the array. In each group corresponding to a distance range, we estimate the average reconstruction error and the corresponding standard deviation. As expected, performance decreases when the interpolation position moves away from the center, although it can be noticed that with methods CS1 and CS2 they decrease slower than with method CS0.

Figure 7 shows the performance of the interpolation when synthetic noise ϵ_n is added to the measurement signals. The x-axis is, on a dB scale, the energy of the additional noise over the energy of the measured signals: $\|\epsilon\|_{\ell_2}/\|s\|_{\ell_2}$. It can

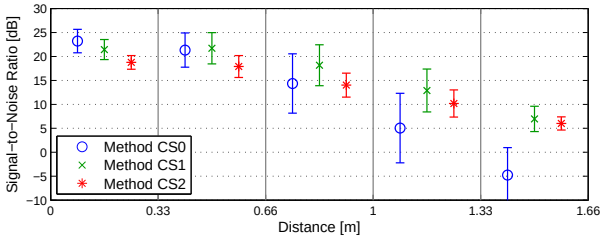


Fig. 6. Evaluation according to the distance from the center of the array. On measured RIRs.

be observed that, as expected, performances decrease when the noise level increases. At high noise levels, method CS2 appears more robust than method CS0 and CS1: the least-square projection of methods CS0 and CS1 tries to fit the whole (noisy) signal with the model, while the "sparse" method CS2, with fewer parameters, intrinsically behaves as a denoising framework.

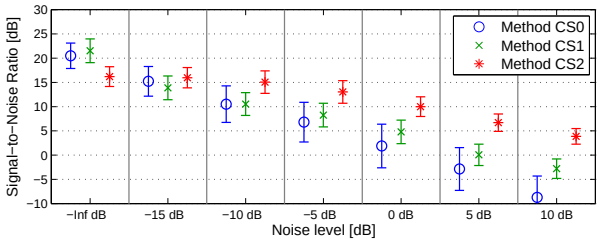


Fig. 7. Evaluation according to the level of the additional noise. On measured RIRs.

An interesting numerical experiment is the comparison of the methods according to the array configuration. In figure 8, we have numerically simulated and tested 6 different array configurations with 125 microphones: 3 random arrays (with a uniform distribution within the cube with side 2m), the experimental array (which is approximately random, cf. fig. 2), a spherical array with radius 1.24m (for which the volume is 8m^3 as the cube), and a regular array (where the receivers are uniformly positioned within the cube with side 2m). The noticeable result is that: as suggested by the RIP for the Compressed Sensing framework, random arrays give best results than regular arrays. For example, methods CS0 and CS1 totally fail for the spherical array. Moreover, the observation of the results of the random arrays (rd1-3 and xp), shows a rather good reproductibility for any random configuration. Further research should investigate arrays with higher performance, with respect to their specific geometry.

As mentioned earlier, when the cutoff frequency f_c increases, the modal density strongly increases, and the sparsity assumption becomes less and less valid. Indeed, the number Q of theoretical modes can be computed (cf. [12]), and table II shows that it increases faster than f_c. As expected, results for methods CS0 and CS1 decrease when f_c increases, cf. fig. 9. Note that stages (a) and (b) of sec. III-C, for CS0 and CS1, cannot be led if the number of available samples N is smaller than the number of estimated modes N_s. For this reason, we need to limit the number N_s for high f_c (cf. tab. II for $f_c = 375$Hz and 400Hz). On the contrary, the number N_a of

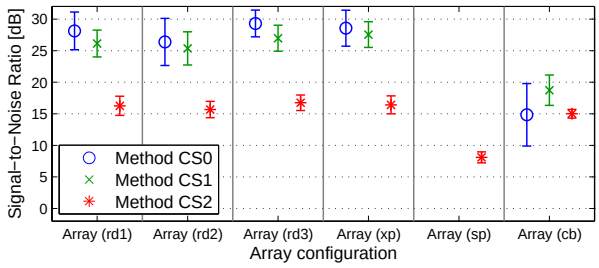


Fig. 8. Numerical evaluation of six different microphone arrays: 3 random arrays (rd1-3), the experimental array (xp), a spherical regular array (sp) and a cubic regular array (cb). The SNRs of methods CS0 and CS1 for the spherical array (sp) are out of range, at almost -12 dB (not displayed). On simulated RIRs.

estimated modes of method CS2 (or equivalently the number of iterations) is not constrained by the number of available samples. Even, we can estimate at more frequencies than the actual number of modes. This is experimentally confirmed on fig. 9, with a remarkable stability for method CS2. Here, only computational issues prevent us from testing at higher f_c, and the cutoff frequency above which method CS2 starts to fail could not be observed here.

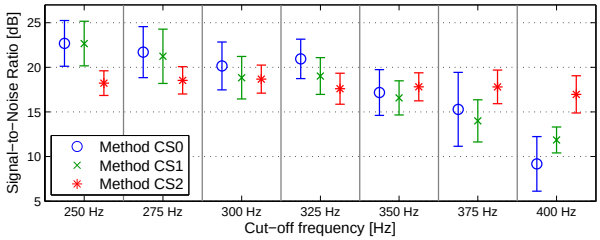


Fig. 9. Results for different cutoff frequencies. On measured RIRs.

f_c [Hz]	250	275	300	325	350	375	400
Q	120	167	221	291	368	465	569
N_s	120	167	221	291	368	426	453
N_a	144	200	265	349	442	558	683
F_s [Hz]	625	694	744	822	868	947	1008

TABLE II
PARAMETERS OF THE EXPERIMENT OF FIG. 9.

Finally, we have checked the robustness of the three methods with respect to the geometry of the room, in particular when the measured room gets further away from the "ideal" empty rectangular room, cf. fig. 10. The acoustics of the room have been significantly changed by opening the windows and the door, and by placing a chair and a large wooden panel. Moreover, the used loudspeaker is directional and we have tested two different orientations. Experimental results for RIR interpolation show that the performance of the three methods was not significantly affected by this change of geometry of the room, or the orientation of the used loudspeaker.

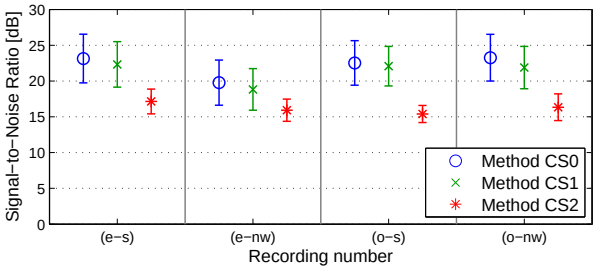


Fig. 10. Experimental tests of other room and loudspeaker configurations. The index (e) means **e**mpty room and (o) means with **o**bstacles, which here means opening the windows and the door, and placing a chair and a wooden panel. The second index defines the direction of the loudspeaker: (s) means **s**outh, the baffled loudspeaker is oriented towards the microphones array, and (nw) means **n**orth-**w**est, the baffled loudspeaker is oriented in an other direction, at 135° from the array. On measured RIRs.

VII. Conclusion

This paper shows that, at low frequencies, the sampling of the full acoustic wavefield in a room is possible with a number of microphones significantly lower than would be required by Shannon-Nyquist sampling theorem. Justified by the modal theory, we have used the Compressed Sensing framework to interpolate the Plenacoustic Function in a 3D-space domain of interest Ω.

The reduction in the number of measurements / microphones allowed by Compressed Sensing can be important in practical applications. However, it comes with a computational cost that can rapidly become prohibitive. The three algorithms presented in this paper have been tuned so that they still can run in reasonable time: the MATLAB analyses of section VI spent almost one hour on a workstation with a 6 core CPU at 3GHz and 24Gb of RAM.

As shown in section VI, the two first algorithms (methods CS0 and CS1) give good results in favorable cases, whereas the last one (method CS2) seems more robust in noisy conditions, and operates on a larger bandwidth. Furthermore, a detailed comparison of methods CS0 and CS1 shows that CS1 is more robust especially with respect to the distance to the center (cf. fig. 6). This observation justifies the use of cross-validation for the selection of the approximation order R (cf. sec. III-C).

In this work, the RIRs reconstruction is limited to the lower frequencies of the spectrum: in this limited part of the spectrum it operates on a time interval that covers the whole duration of the RIRs. A complementary approach can similarly interpolate the early part of the RIRs over a wide frequency range with the same microphone array, using a sparsity assumption of the early reflexions (cf. [24]).

Here, we consider a fixed source at $\vec{X}_s$ and a moving receiver in a domain Ω. Using the reciprocity properties, we get the RIRs for a fixed receiver at $\vec{X}_s$, from a moving source in Ω. Further work should consider both moving source and receiver.

VIII. Acknowledgments

The authors want to thank François Ollivier, Dominique Busquet and Christian Ollivon, from Institut Jean Le Rond d'Alembert - UPMC, for their precious help in the making of the microphone array and in the acquisition of the RIRs.

Appendix

In this section are given some details about the plane wave approximation of section III-B. Let be

$$\Delta u + k^2 u = 0,$$

the Helmholtz equation of solution u with wavenumber k.

Previous studies [5] have shown that, under some conditions on the domain of interest Ω, the following approximation (25) of the solution u as sums of product of spherical harmonics $Y_{\ell,m}$ and spherical Bessel functions j_ℓ, is well behaved.

$$u(\vec{X}) \approx \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} b_{\ell,m} Y_{\ell,m}(\theta,\varphi) j_\ell(k\rho) \qquad (25)$$

in spherical coordinates (ρ,θ,φ). These components can in turn be approximated by sums of plane waves, giving a plane wave approximation of solutions to the Helmholtz equation:

$$u(\vec{X}) \approx \sum_{r=1}^{R} a_r \, \mathrm{e}^{j\vec{k}_r\cdot\vec{X}} \qquad (26)$$

where the wavevectors $\vec{k}_r$ are on the sphere of radius k. Note that this sampling should cover all the sphere, but does not depend on the particular field to be approximated.

These approximations are valid not only in a ball for the spherical harmonics case, or in a box for the plane waves case, but in any domain as long as it is star-convex (it is in particular valid for all convex domains), and are independent on the boundary conditions at the border of the domains. Thus the only condition needed to used approximations (25) and (26) is the star-convexity of the domain of interest Ω (no assumptions are needed on the domain of propagation or on the sources).

Consequently, first, this allows the R-order approximation of the modes ϕ_q of equation (6), second, this validates the observation of the 4D-FT spectrum of the PAF given in section II-C. Moreover, considering the above remarks, the evanescent waves can be also approximated by (6) and the observation of sec. II-C is also valid near the walls and the source.

References

[1] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, October 2006.

[2] C. Masterson, G. Kearney, and F. Boland, "Acoustic impulse response interpolation for multichannel systems using dynamic time warping," in *35th Conference of the Audio Engineering Society*, London, England, February 2009.

[3] Y. Haneda, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 709–717, November 1999.

[4] R. Mignot, G. Chardon, and L. Daudet, "Compressively sampling the plenacoustic function," in *SPIE conference Wavelets and Sparsity XIV*, vol. 8138 813808-1, August 2011, 10 pages.

[5] A. Moiola, R. Hiptmair, and I.Perugia, "Plane wave approximation of homogeneous Helmholtz solutions," *Zeitschrift fr Angewandte Mathematik und Physik (ZAMP)*, vol. 62, pp. 809–837, 2011.

[6] G. Chardon, A. Leblanc, and L. Daudet, "Plate impulse response spatial interpolation with sub-Nyquist sampling," *Journal of Sound and Vibration*, vol. 330, pp. 5678–5689, November 2011.

[7] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, March 1986.

[8] R. Roy and T. Kailath, "Estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[9] G. Chardon and L. Daudet, "Optimal subsampling of multichannel damped sinusoids," in *Proceedings of the 6th IEEE Sensor Array and Multichannel Signal Processing workshop (SAM 2010)*, Israel, October 2010, pp. 25–28.

[10] F. Zotter, "Sampling strategies for acoustic holography/holophony on the sphere," in *NAG-DAGA*, Rotterdam, Netherlands, March 2009, p. 4 pages.

[11] P. Leopardi, "A partition of the unit sphere into regions of equal area and small diameter," *Electronic Transactions on Numerical Analysis*, pp. 309–327.

[12] H. Kuttruff, *Room Acoustics*, 4th ed. Spon press, October 2000.

[13] E. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, March 2008.

[14] R.G.Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, July 2007.

[15] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4206–4215, December 2004.

[16] E. Candès, "Compressive sampling," in *International Congress of Mathematicians*, Madrid, Spain, August 2006, pp. 1433–1452.

[17] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, February 2006.

[18] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, 1st ed. Springer, August 2010, 376 pages.

[19] J. Lagarias, J. A. Reeds, M. Wright, and P. Wright, "Convergence properties of the nelder-mead simplex method in low dimensions," *SIAM Journal of Optimization*, vol. 9, no. 1, pp. 112–147, 1998.

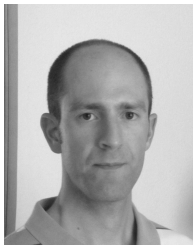
[20] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, April 1976.

[21] P. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1527–1529, November 1986.

[22] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *Journal of the Audio Engineering Society*, vol. 49, no. 6, pp. 443–471, 2001.

[23] N. Ono, H. Kohno, N. Ito, and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'09)*, Mohonk, USA, October 2009, pp. 161–164.

[24] R. Mignot, L. Daudet, and F. Ollivier, "Compressed sensing for acoustic response reconstruction: interpolation of the early part," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'11)*, Mohonk, USA, October 2011, pp. 225–228.



Rémi Mignot received the Dipl. Ing. Degree from Institut Galilée of University Paris XIII, and the Master Degree ATIAM of Pierre & Marie Curie University (UPMC), Paris, France. In 2009, he received a PhD in Signal and Image Processing of Télécom ParisTech with laboratory Analysis/Synthesis Team at IRCAM, France. Then, he did a post-doctoral research in the Langevin Institut (ESPCI ParisTech and UPMC), supported by the Agence Nationale de la Recherche (ANR), project ECHANGE (ANR-08-EMER-006), where he studied the interpolation of Rooms Impulses Responses using Compressed Sensing. He is currently working at Aalto University, Finland, with a Marie Curie post-doctoral fellowship.

Gilles Chardon Gilles Chardon received the engineering degrees of École Polytechnique and Telecom ParisTech in 2009, as well as the MSc ATIAM of Université Pierre et Marie Curie, Paris VI. After working towards his PhD at Institut Langevin, he is now postdoc with the Mathematics and Signal Processing group of the Acoustics Research Institute of the Austrian Academy of Sciences in Vienna. His main research interests include sparse representation of acoustical fields, inverse problems and numerical analysis in acoustics.

Laurent Daudet (M'04-SM'10) studied at the Ecole Normale Supérieure in Paris, where he graduated in statistical and non-linear physics. In 2000, he received a PhD in mathematical modeling from the Université de Provence, Marseille, France. After a Marie Curie post-doctoral fellowship at the C4DM, Queen Mary University of London, UK, he worked as associate professor at UPMC (Paris 6 University) in the Musical Acoustics Lab. He is now Professor at Paris Diderot University - Paris 7, with research at the Langevin Institute for Waves and Images, where he currently holds a joint position with the Institut Universitaire de France. Laurent Daudet serves as associate editor for the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, and is author or co-author of over 120 publications (journal papers or conference proceedings) on various aspects of acoustics and audio signal processing, in particular using sparse representations.